Object Segmentation in Images using EEG Signals

Eva Mohedano[1]
eva.mohedano@insight-centre.org

Graham Healy[1]
ghealy@computing.dcu.ie

Kevin McGuinness[1]
kevin.mcguinness@dcu.ie

Xavier Giró-i-Nieto[2]
xavier.giro@upc.edu

Noel E. O'Connor[1]
noel.oconnor@dcu.ie

Alan F. Smeaton[1]
alan.smeaton@dcu.ie

[1]Insight Centre for Data Analytics
Dublin City University
Glasnevin, Dublin 9, Ireland

[2]Image Processing Group
Universitat Politècnica de Catalunya
C. Jordi Girona, 1-3. 08034 Barcelona

ABSTRACT

This paper explores the potential of brain-computer interfaces in segmenting objects from images. Our approach is centered around designing an effective method for displaying the image parts to the users such that they generate measurable brain reactions. When an image region, specifically a block of pixels, is displayed we estimate the probability of the block containing the object of interest using a score based on EEG activity. After several such blocks are displayed, the resulting probability map is binarized and combined with the GrabCut algorithm to segment the image into object and background regions. This study shows that BCI and simple EEG analysis are useful in locating object boundaries in images.

Categories and Subject Descriptors

H.1.2 [**User/Machine System**]: Human information processing; I.4.6 [**S**]: egmentation; C.3 [**Special-Purpose and Application-Based Systems**]: Signal processing systems

General Terms

Experimentation, Design, Algorithms

Keywords

Brain-computer interfaces, Electroencephalography, rapid serial visual presentation, Object segmentation, Interactive segmentation, GrabCut algorithm

1. INTRODUCTION

The human brain is capable of processing audiovisual information in a fashion that, nowadays, clearly outperforms machines in several applications. The multimedia research community is constantly trying to simulate the brain's behaviour to later leverage its innate computational possibilities through machinery. However, a deep understanding of the human brain remains one of the greatest scientific challenges. Recent initiatives, such us the Human Brain Project in Europe or the the BRAIN Initiative in the United States, have identified its exploration as one of the Grand Challenges of our time.

Although humans consistently outperform computers in the semantic interpretation of multimedia signals [12], the computational and storage power of machines can be scaled and networked dramatically beyond individual human capacities. These two observations are the foundation of the human computational technologies, which exploit the best of both by defining collaborative strategies. The steady decrease in the cost of EEG (Electroencephalography) systems in recent years has made these non-invasive Brain-Computer Interfaces (BCIs) accessible beyond the traditional disciplines that typically availed of this technology [18, 23]. Visual analysis is one such field, with recent publications exploring the potential of EEG signals for image retrieval [10, 26, 25] and object detection [5, 14].

The use of brain-computer interfaces is, however, still limited, primarily because the motor (or speech) capabilities of most humans provide richer interaction methods than BCIs. For this reason, many current applications use BCIs as a secondary interaction source to complement another primary one, or as a tool for scientists to study human behaviour [11]. Brain-computer interfaces, however, have the potential to be enormously beneficial for seriously impaired people, such as those affected by *Locked In Syndrome* (LIS). These individuals are paralysed of nearly all voluntary muscles, so are disabled from motion and speech. Vision is always intact, although in extreme cases even eye movement is restricted [1], in which cases BCIs represent the only opportunity to interact with the world.

Although a controversial discussion topic between neuroscientists, some authors claim to have observed consciousness with EEG devices on patients with persistent vegetative state [7], which may open a door to a certain interaction with them. For these reasons, and as explained in [9], *BCI systems hold great promise for effective basic communication capabilities through machines, e.g. by controlling a spelling program or operating a neuroprosthesis.* The use of EEGs for these type of assistive technologies has been previously explored in applications like letter-by-letter spelling [21] or the control of robots [2, 19].

The objective of this work is to demonstrate that BCI interfaces are useful in tasks beyond spelling out words. We

MM'14, November 3–7, 2014, Orlando, Florida, USA.

http://dx.doi.org/10.1145/2647868.2654896 .

focus here on interaction with multimedia: specifically, object selection and segmentation in images. The capacity to perform such segmentation using a BCI interface potentially has both practical and creative applications, such as selection of specific objects for similarity search, and mixing objects from different sources to create a new composition. We propose a system capable of accurately selecting an object in an image in a manner that is completely hands-free, using only measured signals from an EEG interface. In this way, previous work exploring image retrieval (global image scale) [10, 26, 25] and object detection (coarse local scale) [5, 14] are extended to a pixel-level object segmentation. This task is addressed by applying the human computation paradigm, using noisy EEG signals to seed the well-known GrabCut [22] segmentation algorithm.

This remainder of the paper is structured as follows. Section 2 reviews previous work exploring the use of EEG signals for multimedia analysis. Section 3 provides an overview of the entire system architecture, which is described in detail in Sections 4, 5, and 6. Section 7 presents the results from out experiments. Section 8 gives conclusions and outlines future research directions.

2. RELATED WORK

Previous work combining BCI and computer vision [10, 25, 13] have been focused primarily on image retrieval and object detection. In such work images are presented to participants according to the *oddball paradigm*. This approach consists of presenting a "target" image among many "distractor" images via Rapid Serial Visual Presentation (RSVP) [24]. The presentation rate of the images is high, around 10Hz, so that a specific signature in the corresponding EEG signals is produced when the user observes the target images (or rare stimulus). This signature is known as a P300 wave and it is a kind of Event-Related Potential (ERP) associated to the process of recognising a relevant visual stimulus [15]. The wave's primary characteristic is a positive peak in the EEG signal 300ms after the visual stimulus was observed.

Two previous works describing a BCI system applied to image retrieval and detection were presented by Wang [25] and Healy [10]. In both cases the authors perform RSVP of images from known datasets at 10Hz to detect those images in which a specific object appears. The main difference between them is that in Wang's paper the user is not asked to press any additional button when a target image is seen. Our work differs from these because it focuses on target windows (or regions) instead of target image detection. The most similar work to ours is Bigdely-Shamlo's paper [5], in which satellite images are explored using local windows to detect those containing airplanes. Bigdely-Shamlo's work, however, assumes that the object fits in a single window, while in our contribution objects are partially represented in an unknown number of windows.

3. SYSTEM ARCHITECTURE

We propose a system that aims to both detect and segment an object from an image using the measured brain signals of the user at the moment of observing a specific region. The idea is to transform the measured EEG responses into a map that gives an estimate of how probable it is that a particular region seen by the user contains the target object, and then use this map to seed a segmentation algorithm. The

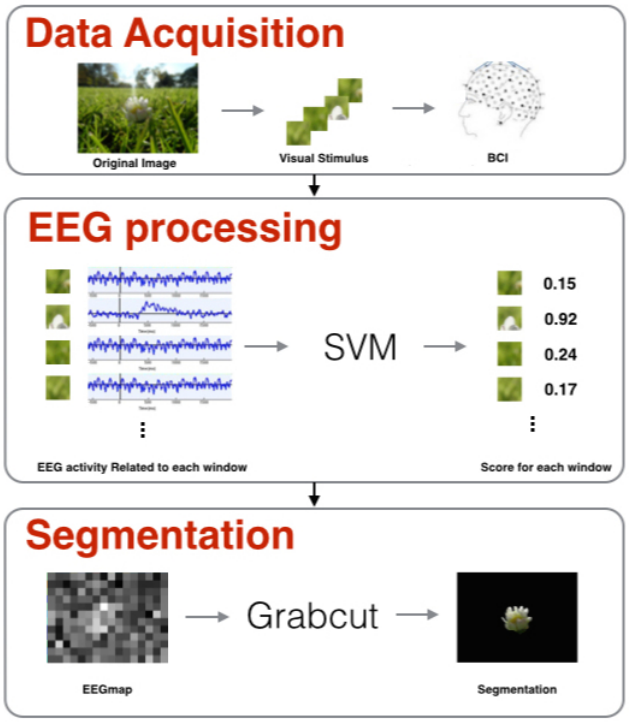

Figure 1: Block diagram of the system.

construction of this map is based on EEG signal classification, as the electrical responses of the brain are known to differ when the user detects a target or rare stimulus in a RSVP scenario.

Figure 1 illustrates the three primary stages of the proposed system:

1. **Data acquisition** (Section 4): in this stage we capture the brain signals related to the visual stimulus.

2. **EEG processing** (Section 5): pre-processing and classification are used to generate the probability maps for the object location. As these maps are built by using EEG analysis, they will be referred to as *EEG maps*.

3. **Segmentation** (Section 6): EEG maps are used to seed the GrabCut object segmentation algorithm [22].

The following sections of the paper describe each stage in more detail.

4. DATA ACQUISITION

This section describes the experimental set-up used to capture the data. First, a new image dataset was created and each image partitioned in blocks of equal size. Each of these blocks are presented at a high rate, in order to generate a measurable response on EEG signals. This stage was validated with a preliminary test with a single user, an important step before starting a larger campaign of data acquisition. After the positive output from the preliminary test, the final experiments reported in the remainder of the paper were based on a population of five people between 21 and 32 years old.

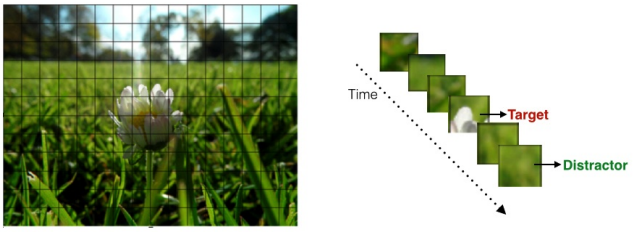

Figure 2: Illustration of RSVP to randomly display different regions of an image.

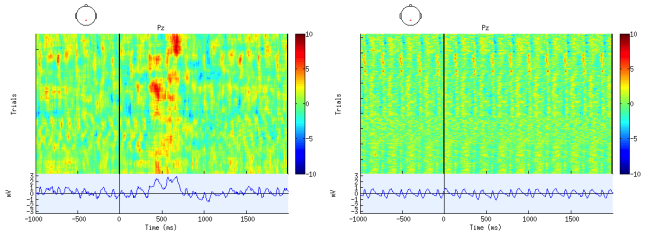

Figure 3: One second before and two seconds after the visual stimulus recorded in the Pz channel for all participants (grand average). Shown are: the amplitudes of the brain waves (top), and the averaged values over all the waves for the target window (bottom-left), and distractor window epochs (bottom-right).

4.1 Image dataset

A novel dataset of 22 images was created to run the experimentation described in this paper. Given the exploratory nature of this work, the images were chosen to include a single object in a background of limited complexity. The dataset includes different configurations regarding the color, shape, and texture of the objects, as well as their relative similarity with the foreground.

The collection consists of 20 new images captured for the purpose of this work and images 38082 and 123074 from the *Berkeley Segmentation Dataset and Benchmark (BSDB)* [16]. The later allow the comparison of the obtained results with other object segmentation approaches. Each of the images has an associated ground truth in the form of a binary mask. In the case of the two BSDS images, the ground truth masks were obtained from a previous work where 100 binary masks from objects where generated from a subset of 96 images [17].

4.2 Windows presentation

The goal of this stage is the generation of the visual stimulus in such a way that they generate a different and measurable cognitive reaction depending on whether they are associated to object or background pixels. The approach adopted is based on the Rapid Serial Visual Presentation (RSVP) [24] of the different windows that compose an image containing an object of interest. The approach follows the same idea described in the papers for image retrieval by using BCI [10, 25, 13] but applied at local scale. This involves partitioning an image into 192 windows and displaying each of them in a fast and random succession (Figure 2). Given the homogeneous scale of the objects in the dataset and the amount of windows, these windows will usually only contain part of the object. In particular, the adopted ratio generated an average of 15% of windows containing parts of the object.

A non-invasive 31 channel BCI with a sample rate of 1kHz was used to capture the brain reaction of the users during the image presentation. The electrodes were located according to the 10-20 system distribution and the experiment was run in a Faraday Cage. This room isolates the participant and equipment to minimize the interference from any other unrelated acoustic or visual events.

Image presentation in the experiments was carried out as follows. First, the entire image was displayed to the participant for five seconds. This allows the user to memorise the visual features of both object and background. Afterwards, the 192 windows of each image were presented at a rate of 5Hz. Each region is shown zoomed and centered on the screen. Preliminary experiments showed participants attention decreased with time. To minimise this effect, we asked participants to count the number of windows containing a part of the object.

4.3 Preliminary experiments

Acquiring EEG data on real users is both laborious and time consuming: in addition to the time required to actually perform the experiments (approximately one hour), it requires scheduling time with volunteers, equipment setup, and precise positioning of the various BCI sensors in a controlled environment. To ensure maximum benefit from each experiment trial, we decided to carry out a set of preliminary small-scale and simulated experiments. The objective of these experiments were: first, to establish whether classification of EEG signals with some reasonable degree of accuracy using our equipment and experiment setup is indeed feasible; second, to determine whether, given a imprecise classification of an EEG signal for a window, it is possible to use this to locate and segment the corresponding object from an image; and third, to guide us in making reasonable choices for the parameters such as the number and size of windows and their presentation rate. We include some details on these experiments here for reproducibility and to justify our design decisions. Positive results at this stage indicated that the system could indeed be effective and helped underpin the full-scale experiments.

4.3.1 Averaging of targets and distractors

The first study focused on the temporal evolution of the EEG signal in those cases where this was captured at the presentation of a target or a distractor window. Given the noisy nature of EEG signals, the observation of any difference between two individual plots from the two classes is challenging. Nevertheless, this noise can be reduced by averaging several signals from the same class and, in this way, distinguish a clear ERP waveform.

Figure 3 compares the same number of target (left) and distractor (right) signals captured in one electrode. The time span goes from one second before the visual stimulus and two seconds after it. The behaviour on the target reactions is different to the distractors, evidencing a peak around 500ms after the stimulus visualization, which is clearly noticed in the averaged waveform across all the erp$target_distractortrials$.

This first result provided the evidence that the adopted RSVP strategy was capable of generating different and measurable brain responses for the two classes of windows. It must be made clear that the future sections in the remainder

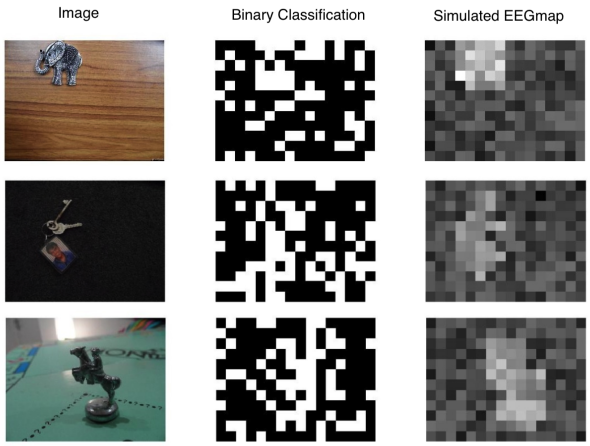

Figure 4: Left: original image; centre: simulation of the window labels (white: windows with object, black: window with background); right: probability map from simulating the SVM scores.

of this paper do not apply any averaging strategy on the EEG signals associated to an image window. All future results presented in late sections (Section 7) are based on the classification the EEG signal obtained with a single trial.

4.3.2 *Binary classification and simulated EEG maps*

A second test was performed to establish the feasibility of distinguishing between target and distractor windows using EEG signals. We posed this as a binary classification problem and trained a binary SVM with RBF kernel classifier with target and distractor EEG signals. 459 EEG signals were used to train the classifier (229 targets and 230 distractors), and 153 EEG signals for testing (76 targets and 77 distractors). The zero-one accuracy obtained was 0.68, which shows sufficient signal is present to achieve better than random classification.

This final preliminary experiment was intended to determine if, given a noisy classification signal from an SVM trained on EEG signals, this could be used to seed a segmentation. We simulated the output of a binary classifier on ground truth images using draws from a Bernoulli distribution with $P(X = 1) = 0.68$ for windows containing a target. Figure 4 (center) illustrates the resulting binary classification maps. The results are, clearly, quite noisy; significant information is lost when the SVM scores are binarized. We therefore chose to use the normalized SVM scores, rather than thresholded decisions, to estimate a soft probability. To simulate SVM scores, we model the distribution of scores given the classification decision as Gaussian, fitting parameters from the data used in the first preliminary experiment, and draw from these Gaussians conditioned on the binary classification decision. Figure 4 (right) shows the resulting generated probability maps, which clearly highlight the object of interest. We chose conditioned Gaussians based on histogram observations; note, however, that this assumption has no bearing on the remainder of the experiments. The simulation results indicate that SVM scores are a useful estimator of the probability that a particular region contains a target.

Table 1: Area Under the Curve (AUC) and Averaged Precision (AP) obtained per user

User	1	2	3	4	5	avg	std
AUC	.63	.75	.73	.78	.65	.71	.06
AP	.22	.33	.30	.45	.23	.31	.08

5. EEG PROCESSING

In this section we describe the actual procedure (i.e. based on what was learnt from the preliminary experiments reported in the previous section) followed to clean and classify the brain signals related to the windows presented to the users. The output generated in this stage are the EEG maps that will be used to produce the final object segmentation for the images.

5.1 Data cleaning and feature vectors

The data was referenced to the Tp9 channel and subsampled from the original 1000Hz rate to 250Hz. For 3 of the users the Tp10 channel was used instead due it was cleaner and, therefore, introduced less noise to the raw signals of the rest of the channels. Then, a band-pass filter from 0.1Hz to 70Hz was applied. By visual inspection, we rejected manually the noisy segments. With the data filtered, we extracted the brain reaction related to the stimulus by selecting one and two seconds pre and post-window presentation (epochs).

For the feature selection, we selected the time region within the epoch that best characterizedd the difference between targets and distractors. As shown in Figure 3, this region is contained between 200ms and 900ms after the visual presentation. The feature vectors are built by concatenating the 31 channels for this time region. The final feature vector is obtained by applying a second subsample to the vectors to reduce the sample rate to 20Hz.

5.2 Binary classification of windows

We worked with the scikit-learn Python library [20] to train the SVM with RBF kernel classifier. The feature vectors were normalized with zero mean and unit standard deviation across each feature component. From the total amount of 22 images, 17 were selected to train the classifier. The EEG vectors related to these images formed an imbalanced set of 435 examples of targets and 2829 examples of distractors, respectively labeled with 1 and 0. An SVM with RBF kernel was trained, and grid search with 5-fold cross validation was used for hyperparameter selection. The parameters selected were the ones that obtained the maximum averaged Area Under the Curve value (AUC) across all the folds.

The final model was tested on 5 images, which contained a set of 130 targets and 830 distractors. Table 1 gives the measured performance.

5.3 EEG maps

The confidence scores provided by the classifier can be graphically represented as an image in the form of *EEG maps*. This score represents the distance that separates the classified sample from the hyperplane [20]. Depending on the sign of this distance, the binary classifier assigns a target or distractor label. The maps are built by normalizing the values assigned to each window between 0 and 1 according

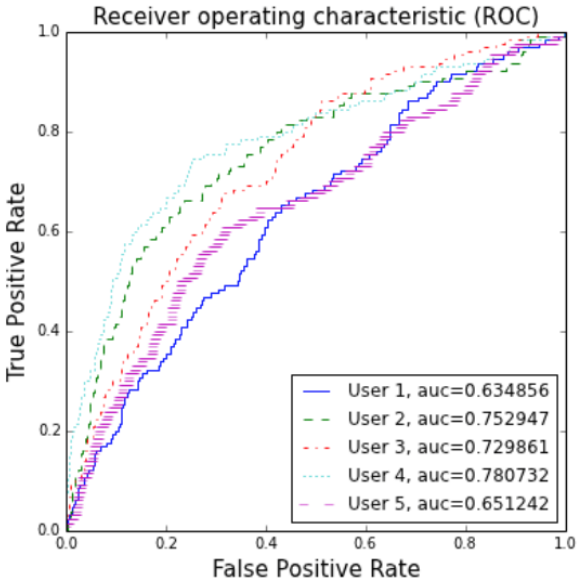

Figure 5: Classification performance per each user in the Reciever Operating Characteristic (ROC) space.

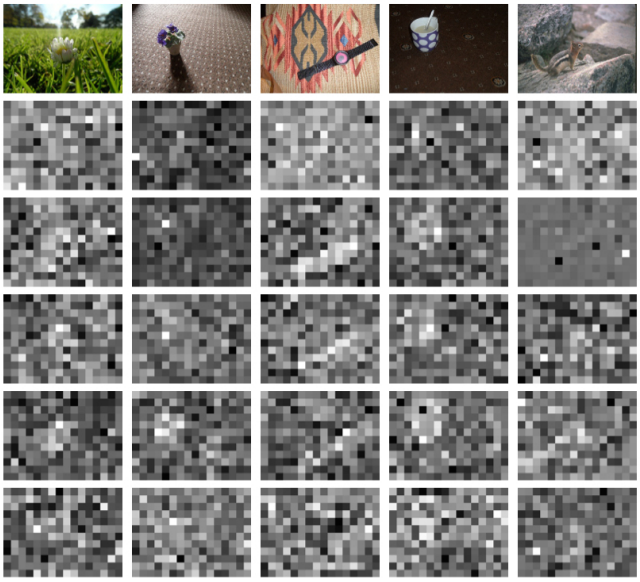

Figure 6: EEG maps for a test set of images. The top row is the original images and the remaining rows are the generated EEG maps for five different participants. Brighter pixels represent higher probabilities.

to:

$$X' = \frac{X - min(X)}{max(X) - min(X)}, \quad (1)$$

where X' represents the EEG map normalized and X the original EEG map.

6. OBJECT SEGMENTATION

The EEG maps constructed in the previous section provide local information about how likely is to find an object part in each window. The final segmentation requires a post-processing of the EEG maps to obtain a pixel-wise binary mask of the object location. Three configurations have been assessed for this task, and for each it was required learn a different set of parameters:

1. Binarization of the EEG maps
2. Filtering and binarization of the EEG maps
3. Filtering and binarization of the EEG maps to seed a segmentation algorithm.

EEG maps are generated after training the SVM model on 17 images. The different values for the segmentation parameters were learned on these training images based on the average performance of the 17 processed EEG maps.

The quality of the segmentation was evaluated with the Jaccard Similarity Index, a popular metric for object segmentation used, for example, in Pascal Visual Object Classes (VOC) Challenge [8]. This measure is made to evaluate the similarity between the final segmentation and ground truth masks. The Jaccard Index has values between 0 an 1, with 1 the maximum similarity between the masks. The measure is defined as the intersection of the two final binary masks divided by the union of both masks:

$$J(A, B) = \frac{A \cap B}{A \cup B}, \quad (2)$$

where A is the segmentation mask and B is the ground truth mask.

6.1 Binarizing the EEG maps

The simplest strategy to quantitatively assess the EEG maps in terms of object localization is to directly convert them into a binary mask. Such binarization is achieved by setting a threshold α, which will consider as targets all those pixels in the EEG map which are higher than α, and label as distractors all the rest. An optimal binarization threshold α_i was estimated for each individual user i by averaging the $\alpha_{i,j}$ values that provided the highest Jaccard index for each training image I_j.

$$\alpha_{i,j} = \underset{\alpha}{\operatorname{argmin}}\, J(M_{i,j}(\alpha), GT_j) \quad (3)$$

where $M_{i,j}$ is the EEG map thresholded by α for user i and image I_j, and GT_j is the ground truth mask for image I_j.

Quantitative results for this approach are presented in Figure 7 have a high density set of windows labelled as target around the object location, especially for user 4.

Table 2 contains the thresholds learned for each of the six users in the test by using 17 images for training. The table also includes the Jaccard index for each user when these thresholds are applying on the 5 test images. The averaged Jaccard index through all the users corresponds to a low 0.14, which points at the poor performance of a direct binarization on the EEG map.

6.2 Filtering and binarization of EEG maps

The binarization approach presented in the previous section presents a first limitation because of the block artefacts introduced by the window boundaries. The window contours do not need to match with the object ones, so in general this

Table 2: Final threshold per user obtained from the EEG maps for training and the final average value obtained applying the threshold on the test set.

User	1	2	3	4	5	avg	std
α	.59	.74	.65	.61	.67	.65	.06
J	.10	.15	.18	.21	.11	.23	.17

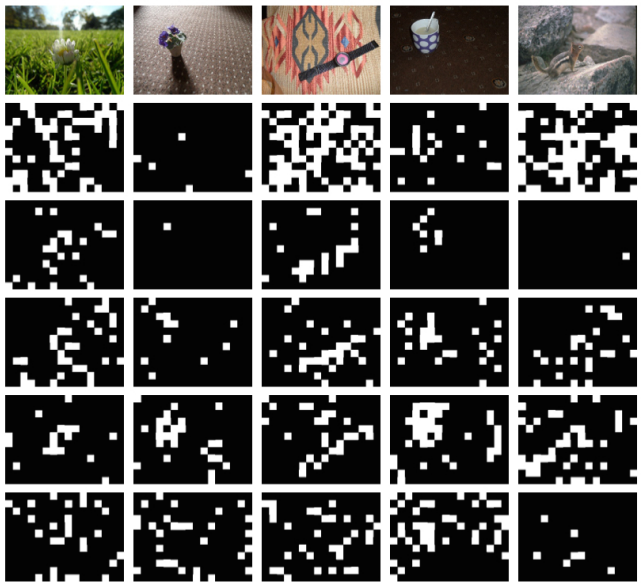

Figure 7: Binary masks after thresholding the EEG maps. The top row is the original images and the remaining rows are the binary masks for five different participants.

lack of resolution is partially responsible of the bad performance of the solution. In addition, the spatial relationship between the windows is completely ignored, without any contextual analysis that may provide coherence to the overall composition.

In this section, a low-pass filter is added before thresholding the maps to reduce block artefacts. With this filter, the isolated false positive windows of the background can be reduced and the high compact windows around the object will mutually reinforce. Equation (4) describes the filter mask (kernel) that is convoluted with the image. The (x, y) values are the horizontal and vertical distances from the origin to a certain point of the kernel. The kernel takes standard deviation σ as a parameter defining the spatial extension of the filter:

$$G(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{x^2+y^2}{2\sigma^2} \qquad (4)$$

The Gaussian filtering and posterior binarization of the resulting EEG map requires defining the two parameters α and σ. As in the previous section, these were selected via minimizing error the training dataset. In this case, though, the Gaussian filtering changes the dynamic range of the EEG maps the threshold, which is no longer between 0 and 1. For this reason, the binarization threshold is not learnt as an absolute value but as a normalised coefficient $p \in [0, 1]$ referred to the dynamic range of the EEG map:

$$\alpha_{i,j}(p) = min(F_{i,j}) + p \cdot (max(F_{i,j}) - min(F_{i,j})), \qquad (5)$$

where $F_{i,j}$ the filtered EEG map of user i for image I_j.

The procedure used for optimisation was to select the parameters (σ, α) that generated the maximum averaged Jaccard Index over all the images of the train set. 70 values for sigma ($\sigma \in [0, 70]$) were tested to filter the EEG map. For each filtered map, 100 different values were tried by varying p from 0 to 1, and the binarization threshold $\alpha_{i,j}$ that maximized the Jaccard was selected, as previously presented in Equation (3). Then, for each image a optimal combination (σ, α) that maximized the Jaccard was obtained. Finally, the parameters used in the test set were set by averaging the 17 pairs of optimal parameters computed for the training set.

Table 3: Averaged percentage (normalized to one) and σ per user obtained from the train set and final Jaccard index obtained on the test set by using these parameters.

User	1	2	3	4	5	avg	std
p	.61	.70	.63	.65	.69	.66	.04
σ	39.24	33.29	26.35	29.18	33.53	32.32	4.89
J	.16	.30	.27	.41	.24	.27	.09

The new binary masks shown in Figure 8 present in many cases a single patch located near the actual position of the object, with a shape which is much more natural than the sparse blocks generated in Figure 7. A quantitative analysis of the results is presented in Table 3 and results in an important gain of 79% when comparing the averaged Jaccard indices of thresholding with our without the Gaussian filtering. The results combining a low-pass filter with thresholding the EEG maps produce a cleaner binary masks which contain a better estimation for the object location (Figure 8). However, these values are still too poor to consider these results an accurate segmentation of the object.

6.3 Seeding the segmentation algorithm

The results obtained in the previous section, based only on EEG data, already provide in many cases a rough estimation of the object location. The configuration explored in this section explores the synergy between BCI data and computer vision algorithms. The EEG maps filtered with a Gaussian kernel are used to seed an object segmentation algorithm that can exploit the spatial dependencies between neighbouring pixels. This way the computer vision algorithm is guided by the user in a noisy and approximate fashion.

The segmentation algorithm used is GrabCut [22]. This technique performs a segmentation of an image based on a rough initial segmentation defined by the user, typically by drawing a box around the target object. The pixels outside the box are initially considered as background and the pixels inside as unknown. The technique models separately the pixels labeled as background and the ones labeled as unknown by using a Gaussian Mixture Model (GMM). The unknown pixels are considered foreground pixels in the first iteration. Then, the two GMMs obtained are used to solve a minimization problem via min-cut and produce a first segmentation of the object. After the initial iteration, with the new labels for background and foreground, GMM are

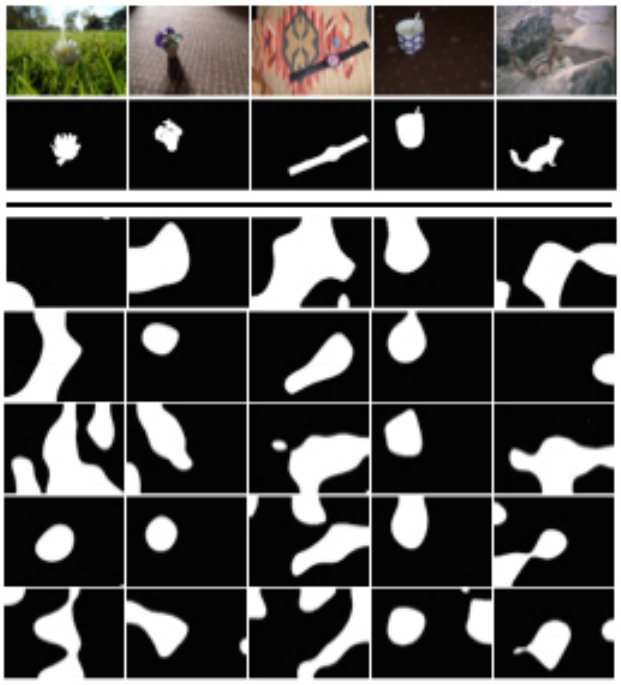

Figure 8: Binary mask after filtering and thresholding the EEG maps. Each row represent the final masks per user.

Figure 9: Final binary segmentations. The images and the ground truth masks are displayed (first an second row). Below, each row represents a different user.

updated and the process is repeated until converge on the final segmentation. Our proposal here is to replace the drawn rectangle by using the EEG maps.

The popular OpenCV [6] implementation of GrabCut was selected for this purpose. The algorithm requires an input a map with pixels marked as: *a*) definitely background; *b*) possible background; *c*) possible foreground; and *d*) definitely foreground. The algorithm requires labels *a* and *b* or *c*. Label *d* is optional, and we have not considered to assign this value due to the noisy nature of the signals.

The initialization of GrabCut with EEG maps requires thresholding the Gaussian filtered map by applying two thresholds: α_1 to separate pixels (a) from the rest, and α_2 to separate the pixels labeled as (b) from pixels of (c). Both thresholds are defined as relative percentages (p_1, p_2), as in the Section 6.2 eq. (5) for the same reason: after applying the low pass filter, the EEG map is unnormalized.

The optimization is realized by randomized search [3], trying combinations of (p_1, p_2, σ) and computing the final Jaccard Index. We used the hyperopt python package [4] for the optimization problem.

The function to optimize (eq. (6)) is the one that computes the average of all the Jaccard Index for the training set given a parameters combination. The optimization is to find the parameters that minimize the error on the averaged Jaccard index:

$$E(p_1, p_2, \sigma) = 1 - \frac{\sum_{j=1}^{N} J(M_{i,j}(p_1, p_2, \sigma), GT_j)}{N}, \qquad (6)$$

where $N = 17$ is the number of training images. First, the EEG maps are filtered by applying a Gaussian filter with the sigma parameter $\sigma \in [0, 70]$, then the filtered map is thresholded at two levels by applying two thresholds: $p_1 \in [0, 0.5)$ and $p_2 \in [0.5, 1.0]$. We randomly pick 1,000 combinations.

Table 4: Optimal parameters (p_1, p_2, σ) per user. p_1, p_2 are the percentages normalized to 1. Accuracy is the final Jaccard index on the test set.

User	1	2	3	4	5	avg	std
t_1	.04	.15	0.02	.27	.17	.17	.18
t_2	.77	.76	.64	.81	.69	.73	.07
σ	35.65	29.56	37.53	3.18	19.10	25.00	14.16
Jacc.	.28	.62	.31	.36	.69	.45	.19

7. RESULTS

As the number of images for testing the system is limited, a cross-validation is performed by switching the images on the test and training set 5 times and producing that way the segmentation of all the dataset. That means that 5 different systems are generated following the pipeline described (Figure 10), where the 5 testing images are always independent from the training.

The results obtained are plotted in Figure 11 for the three strategies to produce the final binary mask introduced in Section 6. Averaged Jaccard accuracies indicate that the configuration of using GrabCut with the filtered and thresholded EEG maps performs better than the other configurations, producing a good binary mask in many of the images. However, in other images, the Jaccard Index is not high enough and the segmentation is noisy.

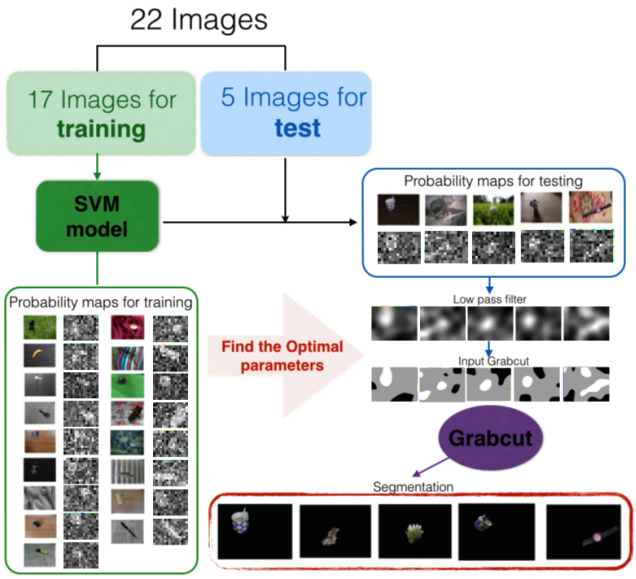

Figure 10: Diagram of the segmentation procedure. The training set is used to generate the SVM model to produce the EEG maps. The same SVM is used on the training images to produce a EEG maps for training, from where the system parameters are estimated. The separated test images ares classified by the SVM and the final segmentation is performed by using the learned parameters.

Table 5: Final Jaccard for each of the five iterations of the cross-validation of the system

	Jaccard A	Jaccard B	Jaccard C
avg	.13	.21	.47
std	.03	.04	.12

Figure 12 presents the visual segmentation for five examples, as well as the intermediate stages. The first three results offer a good qualitative segmentation, while the two last do not succeed in the task. These two failing examples share the characteristic of a very similar distribution between the object and the background. While the EEG map offers a reasonable quality, the GrabCut algorithm fails in the segmentation. This effect is possibly due to the color-driven approach adopted by GrabCut, which basically models foreground and background with color GMM.

It is possible to see that the filtered EEG maps produce a good estimation of the object location that in three of the five examples produce a good segmentation. The segmentation is less accurate in two images, although the location in the processed EEG map is reasonable.

These results show that it is possible to successfully classify the brain reaction produced to detect different parts of a target object, and to produce useful information based on the EEG waves to locate the target object on the images.

7.1 Combining EEG maps of different users

To reduce the noise of the EEG maps, we compute an unique map per image by averaging the EEG maps of the

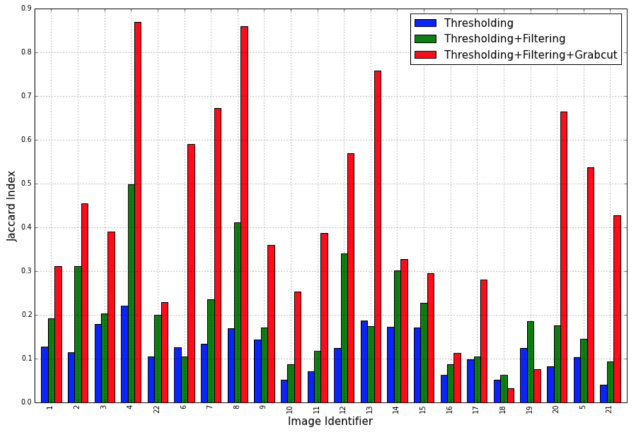

Figure 11: Average Jaccard index across users per image.

different users. The final segmentation is performed following the approach described in the section 6.3. The parameters are picked by averaging across iterations and users ($p_1 = 0.23$, $p_2 = 0.74$, $\sigma = 24.10$).

Qualitative results of the averaged EEG maps provide evidence that combining the individual maps of different users it is possible to generate cleaner EEG maps (Figure 13). The final Jaccard combining the EEG maps of the users outperform in 18 of the 22 images, getting an averaged Jaccard of 0.72, 1.6 times superior to the global result obtained before (Table 13).

7.2 EEG vs. mouse-based interfaces

The analysis of the proposed BCI-based solution is compared with a state-of-the-art solution using a mouse instead of the human-computer interface. The study is based on the two images from the Berkeley Segmentation DataSet (BSDS) [16] we used, which were also considered in a previous study on mouse-based segmentation tools [17]. In that work, four different segmentation techniques were compared in an interactive set up where users draw scribbles to seed the algorithm. That experimentation measured the evolution of Jaccard index with respect to the amount of time that a user was engaged in the operation of the tool.

For the sake of a fair comparison, only the image processing algorithm referred as Interactive Graph Cuts (IGC) has been considered because it is the most similar to the GrabCut solution adopted in our paper. In terms of time evolution, we selected the Jaccard index obtained 45 seconds after the start of the interactive segmentation because this is the closest time stamp to the total display time of an image in our EEG-based system: 43.4 seconds distributed in an initial display of the full image during 5 seconds and 38.4 seconds for the RSVP of 192 windows at a rate of 5 Hz.

The average Jaccard indexes for the EEG- and mouse-based segmentations are presented in the Table 6. The obtained figures clearly show that mouse-based interaction outperforms the proposed EEG-based method which, in addition, requires the costly task of installing the BCI on the user. Note that the high variability on the standard deviation associated to the averaged Jaccard evidence a high variability in the users performance for the EEG results, different to the mouse-based interface, where all the users perform similar.

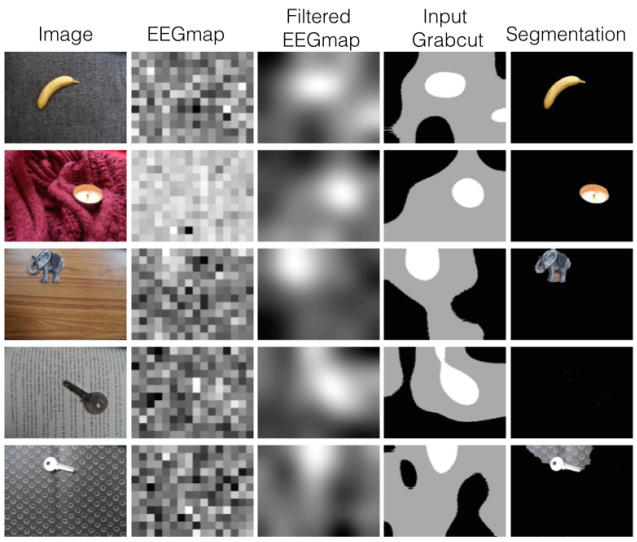

Figure 12: All steps to produce the segmentation combining the EEG maps with GrabCut. Results are from the user 2; segmented images are the set used on the second iteration of the cross-validation.

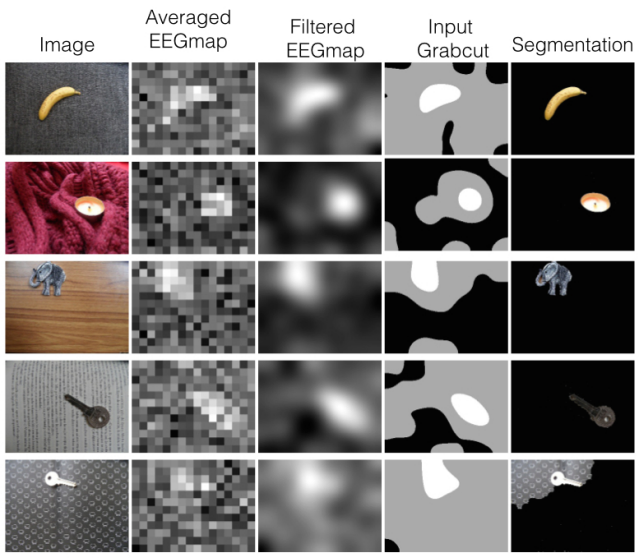

Figure 13: All steps to produce the segmentation combining the EEG maps with GrabCut. The EEG maps are the average of the all the user's maps generated.

8. CONCLUSIONS AND FUTURE WORK

We proposed a system for object segmentation using brain signals. The system is posed as a proof of concept, with the objective being to determine if such a system is feasible.

We designed a specific method of presenting images to associate each image region with its visual brain reaction. Our use of non-overlapping blocks limits the resolution of the generated EEG maps; future work will consider overlapping windows increase the spatial resolution.

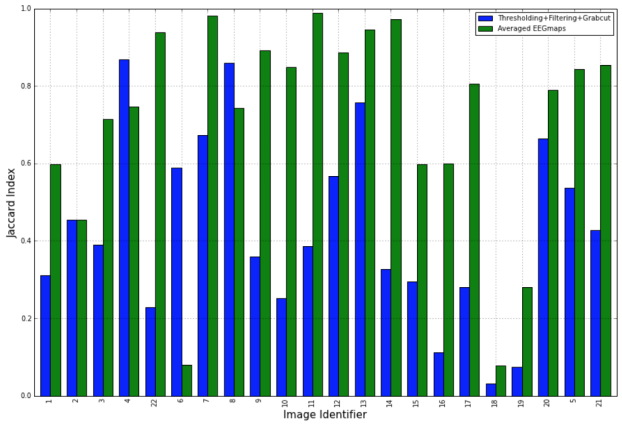

Figure 14: Final Jaccard per image. In blue, EEG maps combined with GrabCut. In green, averaged EEG maps across users combined with GrabCut.

Table 6: Jaccard Index for EEG and mouse-based interaction methods.

	BSDS 38082	BSDS 123074
EEG	$.36 \pm .33$	$.23 \pm .23$
Mouse	$.74 \pm .06$	$.89 \pm .02$

The EEG processing performed in the paper (Section 5) is based on low-pass filtering the EEGdata, epoching the data to identify each window with its brain reaction, training an SVM with the the down-sampled signal, and concatenating EEG channels to form a feature vector. This simple processing gives an AUC of .71; more sophisticated analysis, e.g. Independent Component Analysis (ICA) for channel selection and artifact removal, may improve classifier performance. Better EEG features, such as wavelet features, may also improve classification and, consequently, the quality of the EEG maps.

In Section 6 we investigated three different configurations to produce binary masks. The EEG maps obtained are noisy and require post-processing; a Gaussian low-pass filter was effective in reducing the effect of noise and improves Jaccard accuracy (6.2). Using other filters or morphological operators on the binarized EEG map may improve results. Subsection 6.3 discusses our preprocessing of the EEG maps to set the initial inputs to GrabCut. Future work will consider using the values of the EEG maps directly to set initial terminal capacities of the min-cut graph.

We have shown that the fusion of different user's EEG maps helps to reduce the noise on the probability masks and perform a segmentation that, once combined with GrabCut, notably outperforms the segmentation performance acquired.

Our system shows that it is possible to roughly locate and delineate an object in an image using EEG data, but is far from get the quality on the segmentation of other state-of the art interactive segmentation tools. Nonetheless, this proof of concept opens the door to new interaction modes which may become specially valuable for those people affected by Locked in Syndrome. For them, this work may represent a promising direction in improving their communication for applications such as object selection. If the accuracy of BCI

keeps increasing, and their cost decreasing, it is expected that new applications will appear for this novel human computer interface that has raised the interest in different fields of the multimedia community.

9. ACKNOWLEDGMENTS

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289 and partially funded by the Project TEC2013-43935-R BigGraph of the Spanish Government.

10. REFERENCES

[1] G. Bauer, F. Gerstenbrand, and E. Rumpl. Varieties of the locked-in syndrome. *Journal of Neurology*, 221(2):77–91, 1979.

[2] C. J. Bell, P. Shenoy, R. Chalodhorn, and R. Rao. Control of a humanoid robot by a noninvasive brain computer interface in humans. *Journal of Neural Engineering*, 16(5):432–441, 2008.

[3] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.*, 13:281–305, Feb. 2012.

[4] J. Bergstra, D. Yamins, and D. D. Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *ICML (1)*, volume 28 of *JMLR Proceedings*, pages 115–123. JMLR.org, 2013.

[5] N. Bigdely-Shamlo, A. Vankov, R. Ramirez, and S. Makeig. Brain activity-based image classification from rapid serial visual presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5):432–441, 2008.

[6] G. Bradski. *Dr. Dobb's Journal of Software Tools*, 2000.

[7] D. Cruse, S. Chennu, C. Chatelle, T. A. Bekinschtein, D. Fernández-Espejo, J. D. Pickard, S. Laureys, and A. M. Owen. Bedside detection of awareness in the vegetative state: a cohort study. *The Lancet*, 378(9809):2088–2094, 2012.

[8] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The Pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[9] D. Fernandez-Canellas. Modeling the temporal dependency of brain responses to rapidly presented stimuli in erp based bci. Master's thesis, Northeastern University, 2013.

[10] G. Healy and A. F. Smeaton. Optimising the number of channels in eeg-augmented image search. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, BCS-HCI, pages 157–162, 2011.

[11] R. Hebbalaguppe, K. McGuinness, J. Kuklyte, G. Healy, N. O. Connor, and A. Smeaton. How Interaction Methods Affect Image Segmentation : User Experience in the Task. In *Proc. The 1st IEEE Workshop on User-Centred Computer Vision (UCCV)*, 2013.

[12] X. Hu, K. Li, J. Han, X. Hua, L. Guo, and T. Liu. Bridging the semantic gap via functional brain imaging. *Multimedia, IEEE Transactions on*, 14(2):314–325, 2012.

[13] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, and K. E. Hild, II. A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomputing*, 74(12-13):2041–2051, June 2011.

[14] A. Kapoor, P. Shenoy, and D. Tan. Combining brain computer interfaces with vision for object categorization. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[15] S. J. Luck. *An introduction to the event-related potential technique*. MIT Press, 2005.

[16] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, volume 2, pages 416–423, July 2001.

[17] K. McGuinness and N. E. O'Connor. A comparative evaluation of interactive segmentation algorithms. *Pattern Recogn.*, 43(2):434–444, Feb. 2010.

[18] S. Motomura, Y. Ojima, and N. Zhong. Eeg/erp meets act-r: A case study for investigating human computation mechanism. In N. Zhong, K. Li, S. Lu, and L. Chen, editors, *Brain Informatics*, volume 5819 of *Lecture Notes in Computer Science*, pages 63–73. 2009.

[19] I. Pathirage, K. Khokar, E. Klay, R. Alqasemi, and R. Dubey. A vision based p300 brain computer interface for grasping using a wheelchair-mounted robotic arm. In *2013 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, pages 188–193, July 2013.

[20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[21] B. Roark, B. Oken, F.-O. M., U. Orhan, and D. Erdogmus. Offline analysis of context contribution to erp-based typing bci performance. *Journal of Neural Engineering*, 10(6):432–441, 2013.

[22] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics*, 23(3):309–314, August 2004.

[23] P. Sajda, E. Pohlmeyer, J. Wang, L. C. Parra, C. Christoforou, J. Dmochowski, B. Hanna, C. Bahlmann, M. K. Singh, and S.-F. Chang. In a blink of an eye and a switch of a transistor: cortically coupled computer vision. *Proceedings of the IEEE*, 98(3):462–478, 2010.

[24] R. Spence. Rapid, Serial and Visual: a presentation technique with potential. *Information Visualization*, 1(1):13–19, 2002.

[25] J. Wang, E. Pohlmeyer, B. Hanna, Y.-G. Jiang, P. Sajda, and S.-F. Chang. Brain state decoding for rapid image retrieval. In *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09, pages 945–954, 2009.

[26] A. Yazdani, J.-M. Vesin, D. Izzo, C. Ampatzis, and T. Ebrahimi. Implicit retrieval of salient images using brain computer interface. In *ICIP*, pages 3169–3172, 2010.